THE NETFLIX STORY: REAL TIME MOVIE STREAMING INSIGHTS

REPORT BY: AKUA AGYARE

JULY 2024

INTRODUCTION

This document provides an in-depth insight into my Netflix movie streaming experience. Movies and TV series have always been my topmost source of entertainment and Netflix is my number one stop shop. Netflix is among the popular media and video streaming platforms. They have over 10,000 movies/TV shows on their platform as of 2021. As a Netflix enthusiast with a high affinity for data analytics and storytelling, I carried out this activity to gain deeper understanding of my Netflix viewing habits, ratings and content characteristics. Insights generated will offer personalized recommendations for streaming movies in future.

Objective

Leverage the potential of data analytics and engineering techniques to analyze my Netflix movie streaming data to understand my Netflix movie preferences based on viewing history, ratings and content attributes.

Key tasks/Deliverables

- ❖ Perform ETL (Extract, Transform &Load) processing to analyze trends in viewership habits and content preferences.
- ❖ Build an analytical dashboard including a story report on insights gained and recommendations to Netflix for movie watching in future.

DATA COLLECTION AND DESCRIPTION

Two datasets used were extracted from the Netflix user database (in excel format) for analysis and generating insights. The first has **4 columns** and **1,932 rows** consisting of details of movies/series watched from **2022 to 2023** – movie id, movie title, date viewed and time viewed. The other data set has 1932 rows and 12 columns providing characteristics (genre, producers, year produced, date released, runtime, gross income, languages, votes & ratings) of the movies watched in 2022 – 2023.

APPROACH

Figure 1: Data Flow Diagram

This activity involved collecting data from the **Netflix user database**. The data sets retrieved were converted from excel to csv format using **MS Excel**. Extract, Transform and Load processes were used to ingest all data into **python** for transformation (Data profiling, cleaning, type conversion& feature engineering), and loading into **MySQL database** for querying. The transformed data was imported from MySQL database into **Power BI** for further analysis of viewing habits and content preferences of movies watched in **2023**, resulting in a design of an insightful storyboard.

SUMMARY OF KEY INSIGHTS AND LESSONS

- From the findings, a total of 1.008K movies comprising of 13 genres were watched within 15K days in 2023.

- It can be inferred that, Drama (287), Comedy (142), Action& Thriller (107 each) are the 4 most preferred movie genres.

- Based on the daily trends in viewership, insights indicate that Saturdays, Sundays, Thursdays and Fridays are the best days for watching movies, as they show the highest average runtime per movies watched.

- On a monthly basis, January, February, April & May had the most number of movies watched across all days of the week. This suggests they are the less busy seasons.

- Based on the highest popularity rating and votes, movies produced by Warner Bros, Walt Disney, Twentieth Century Fox, Paramount Pictures & Columbia Pictures would be the main focus in future.
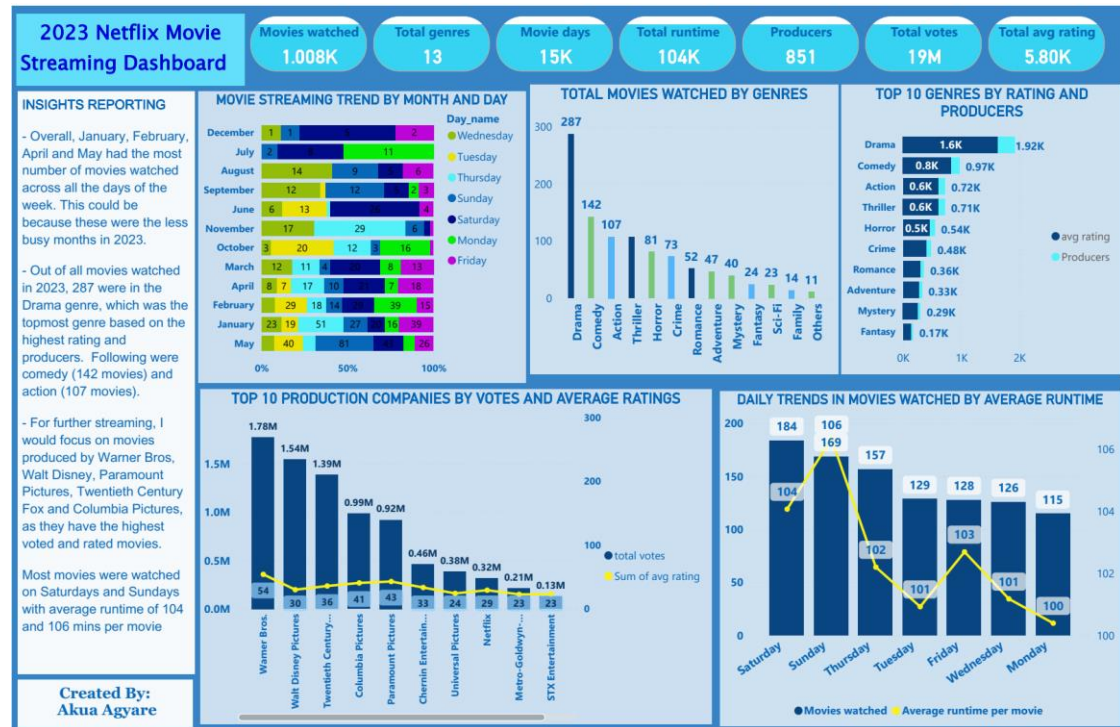

Figure 2: Netflix Movie Streaming Data Storyboard